

Plánování, organizace a hodnocení klinických studií – seminář

MZKPL0311s – týden 2

Tomáš Pavlík

Obsah přednášky

- Klinické studie
- Fáze klinického hodnocení léčiv
- Základní pojmy
- Úvod do analýzy dat v klinických studiích
- Cílové parametry hodnocení
- Hypotézy v klinických studiích
- Experimentální design klinických studií
- Randomizace, zaslepení a odhad velikosti vzorku
- Interim analýza, subgroup analýzy
- Legislativa



týden 1

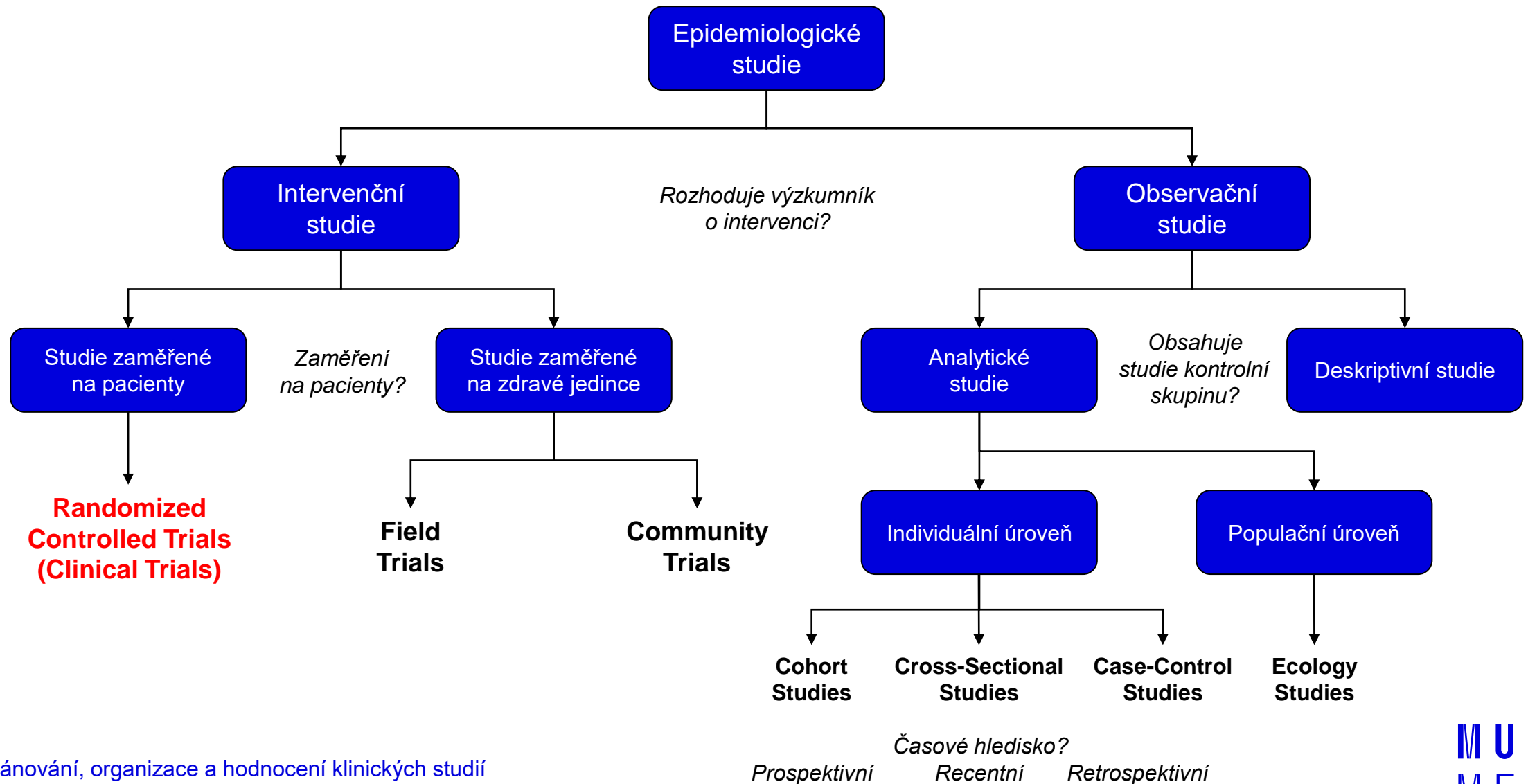
týden 2

Experimentální design klinických studií

Design studie

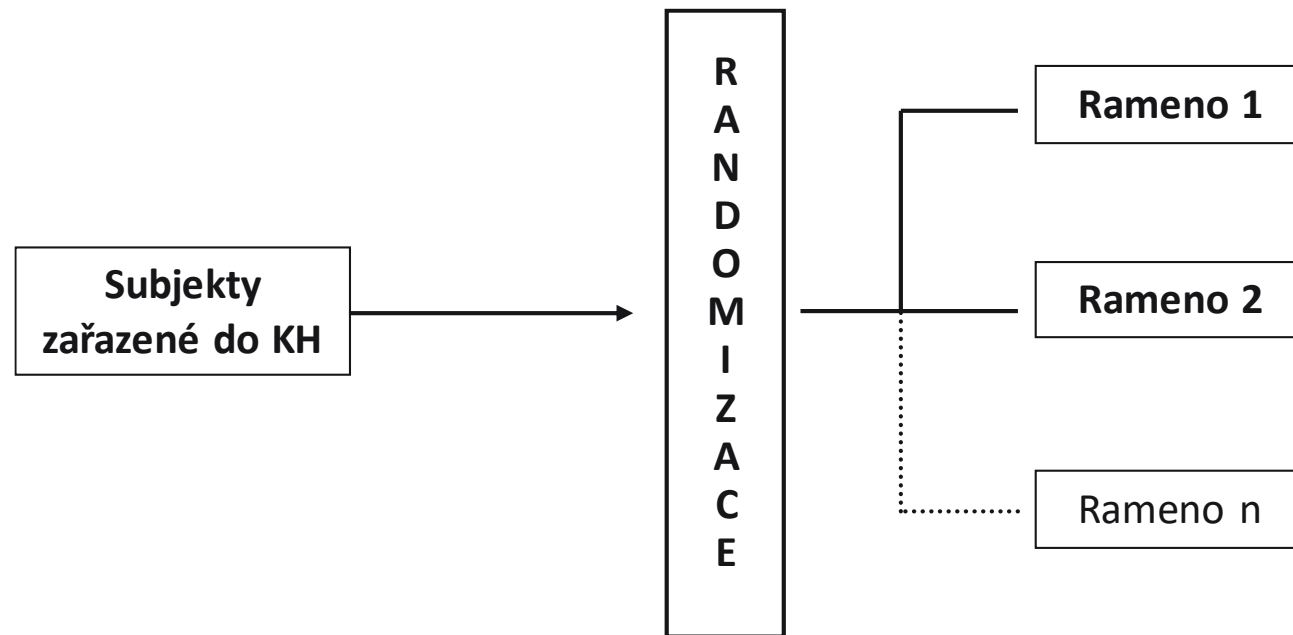
- Základní nastavení studie, adekvátně zohledňuje stanovené cíle
- Určuje jak provedení, tak vyhodnocení (vyhodnocení je tedy s designem bezprostředně svázáno)
- Špatný design většinou vede k tomu, že studii nelze vyhodnotit
- **Epidemiologické studie, neintervenční studie (NIS)**
 - Retrospektivní vs. prospektivní
 - Studie případů a kontrol, průřezové studie
 - Kohortové studie
- **Klinická hodnocení**
 - Vždy prospektivní, intervence (resp. diagnostika, monitoring)

Pozice klinických studií v epidemiologii



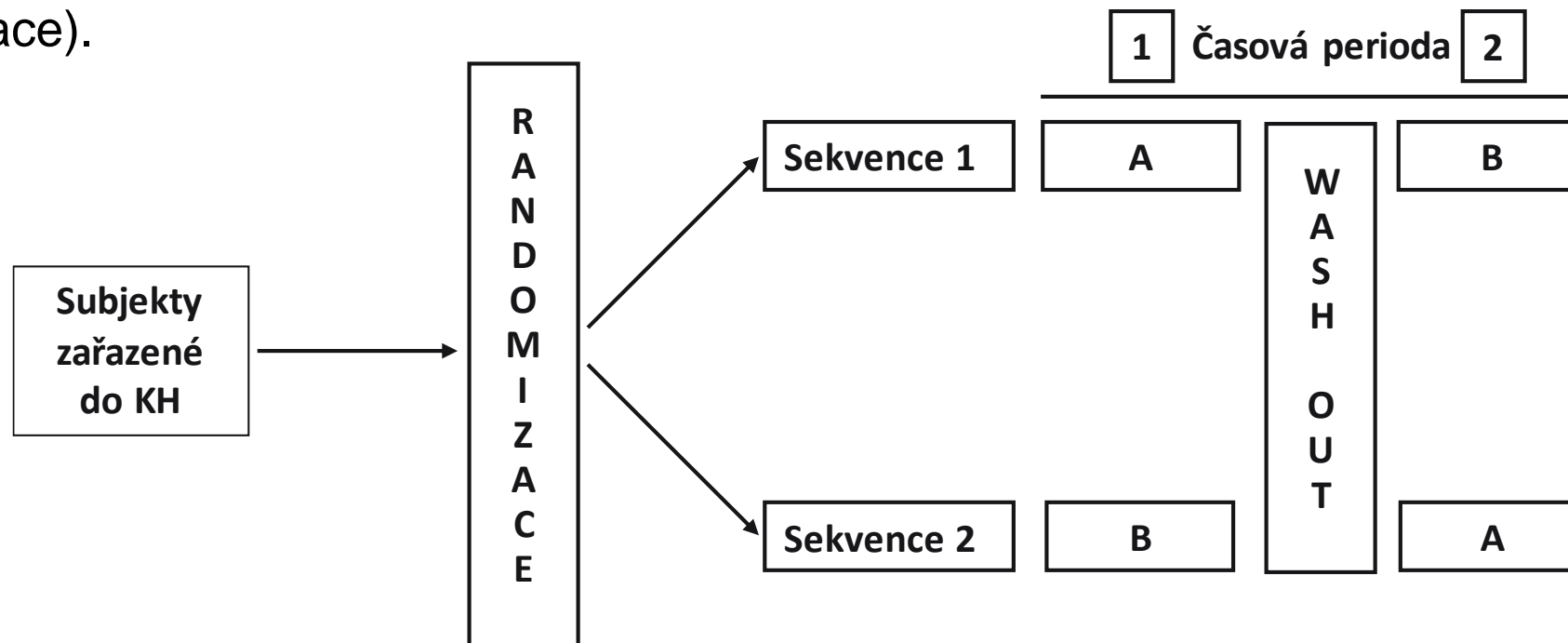
Paralelní uspořádání studie

- Srovnávána účinnost a bezpečnost dvou nebo více intervencí
- Minimalizuje výběrové zkreslení
- “Zlatý standard”
- Studie fáze III (II)



Cross-over uspořádání studie

- Rovněž randomizovaná, kontrolovaná, studie fáze III (II), bioekvivalenční studie (generika).
- Pacient je “sám sobě kontrolou” (pro stejnou přesnost je nutný menší celkový počet pacientů, interpretace).



Cross-over studie II

– Efekt

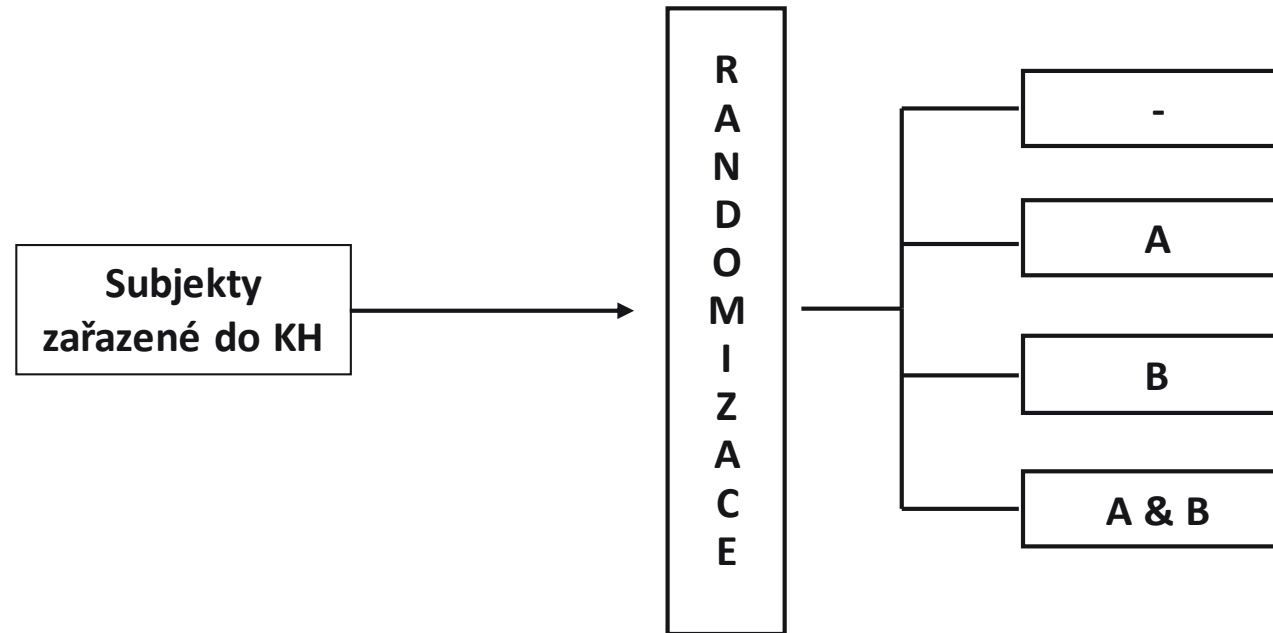
- Intervence
- **Carry-over** (farmakologický, psychologický)
- Vliv časového období (změna podmínek)
- Plus interakce léčby a časového období

– **Není možné realizovat vždy**

- vhodné pro relativně stabilní chronická onemocnění (DM, hypertenze)
- nevhodné pro měnící se onemocnění (onkologie - progrese, vliv sekvence) nebo podmínky (alergie)

Faktoriální design

- Rovněž randomizovaná, kontrolovaná
- Studie spíše “optimalizující” již zavedené intervence



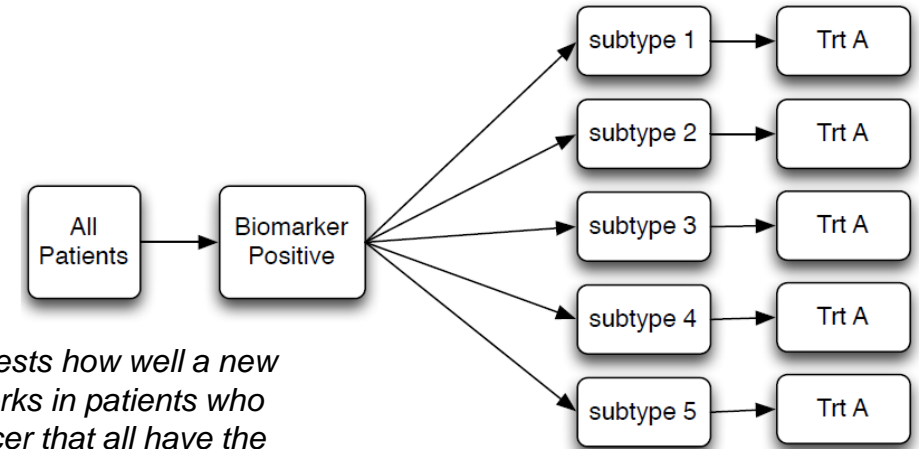
Designy pro hodnocení biomarkerů

„Basket“ a „Umbrella“ designy

(Basket trial, Umbrella trial)

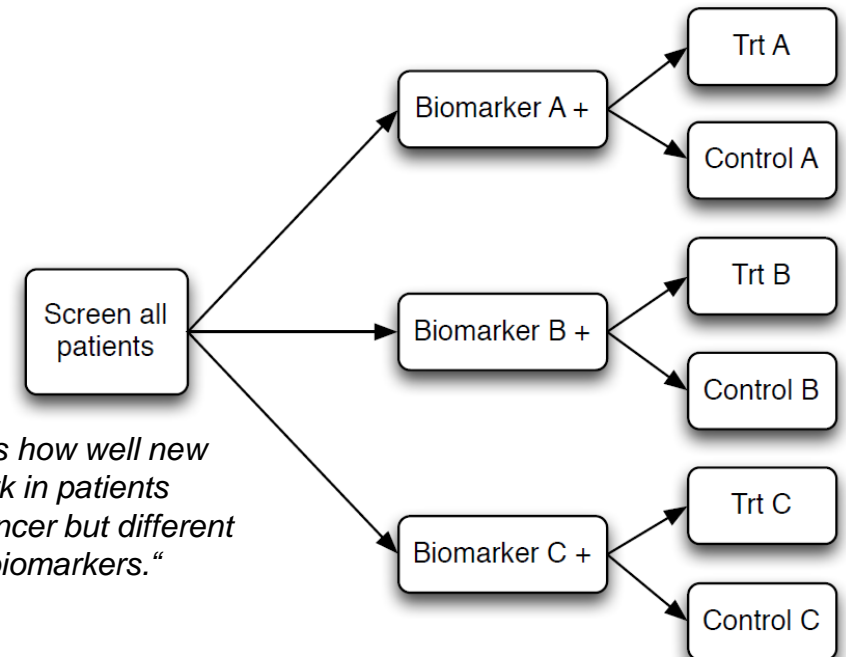
- Mohou být rovněž randomizované a kontrolované
- Studie hledající výhodné kombinace biomarker x intervence
- Typicky studie fáze II

Basket trial:



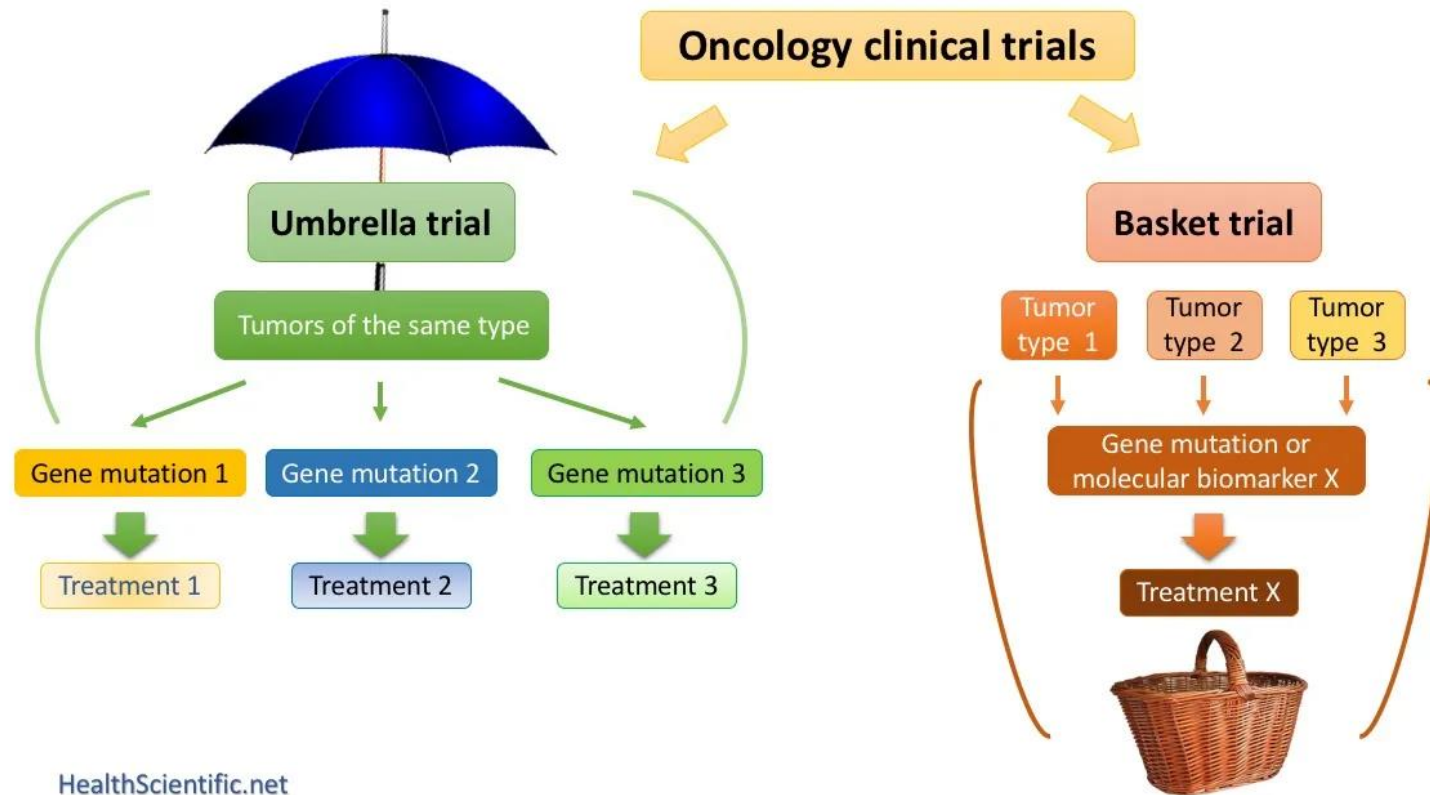
„A type of clinical trial that tests how well a new drug or other substance works in patients who have different types of cancer that all have the same mutation or biomarker.“

Umbrella trial:

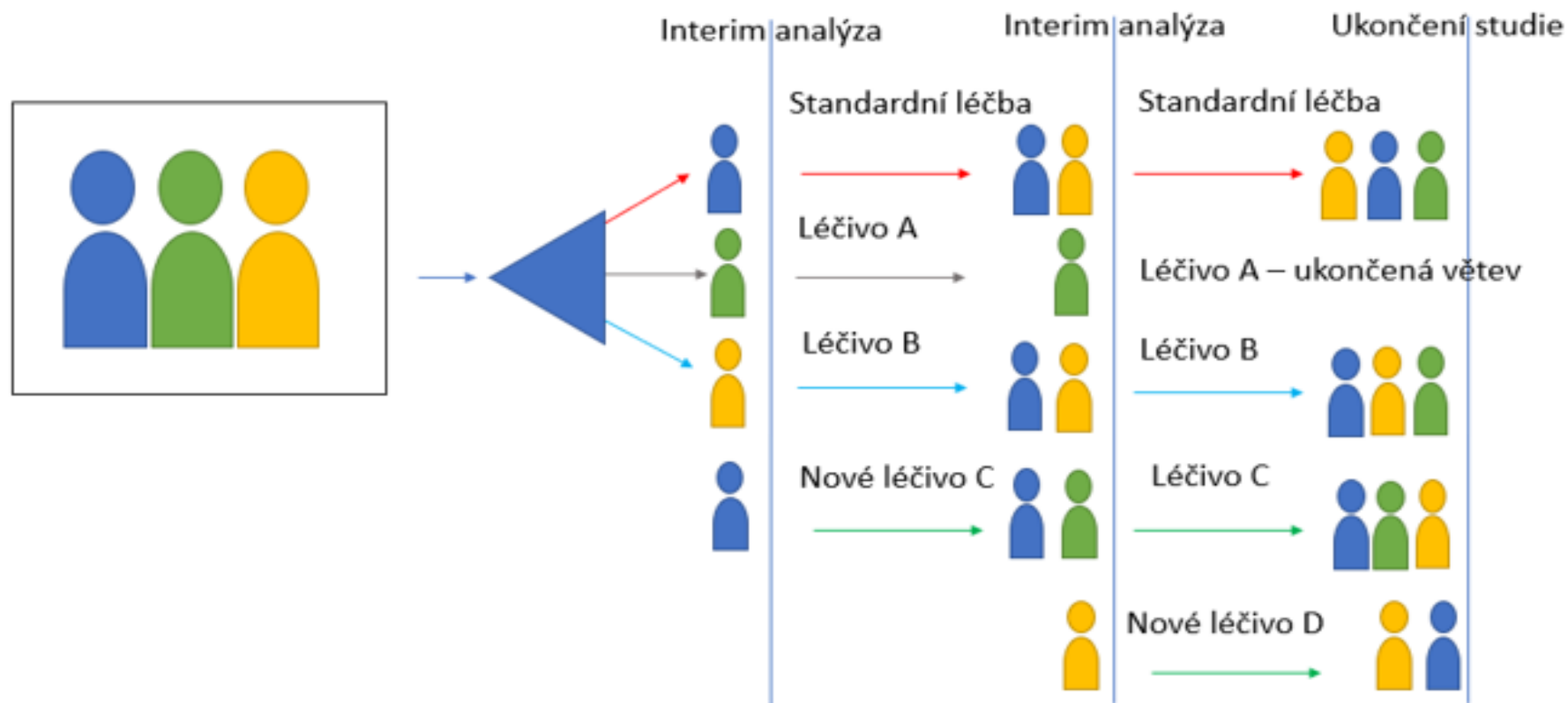


„A type of clinical trial that tests how well new drugs or other substances work in patients who have the same type of cancer but different gene mutations (changes) or biomarkers.“

Basket trial vs. Umbrella trial



Adaptivní design studie



Dose-finding, dose-ranging, dose-response

- Studie fáze I, II
- Stanovení optimální dávky (max. účinek, min. nežádoucí účinek)
- Odpověď vs. dávka
- Zpravidla více dávek (rané fáze více, pozdější fáze srovnávají už jen několik vybraných)
 - Paralelně více dávek, pacienti randomizováni (nutná znalost bezpečnostního profilu)
 - Cross-over
 - Titrace pro každého pacienta (chronická onemocnění)
 - Eskalace dávky

Součást informací o designu studie

- Popis ramen (skupin)
- Přiřazení intervence (randomizace, dle zařazení) zaslepení
- Popis fází
 - screening
 - run-in
 - průběh a ukončení intervence
 - průběh a ukončení následného sledování (follow-up)
- Počet pacientů
- Popis základního sledování během fází (co se bude sledovat, jak často)
- Ukončení účasti pacienta ve studii, ukončení studie

Proč se definují “analyzované populace”

- **Ideální stav:** všichni pacienti zařazení do studie, randomizováni, podstoupí intervenci dle protokolu, ukončí dle protokolu, apod.
- **Realistický stav:** zařazení pacientů nebo průběh intervence neodpovídá protokolu, pacienti ukončí intervenci/studii z bezpečnostních/jiných důvodů, bez udání důvodů, “lost to follow up”, chyby, ...

Mají takoví pacienti být zařazení do vyhodnocení?

- **ANO: může vést ke zkreslení** (např. pokud pacient nedostal léčbu v plné dávce, nebude účinek stejný, jako kdyby ji dostal)
 - **NE: může vést ke zkreslení** (např. tímto z jednoho ramene vyloučení pacienti s horší prognózou, to povede k umělému rozdílu mezi rameny)
- Je nezbytná definice analyzovaných populací (záleží na otázce, která se má zodpovědět) (“analysis population”, “analysis set”)
- Neplést si se studijní populací (“study population”, “patient population”), kterou mají zařazení pacienti reprezentovat (daná indikace, kritéria pro zařazení)

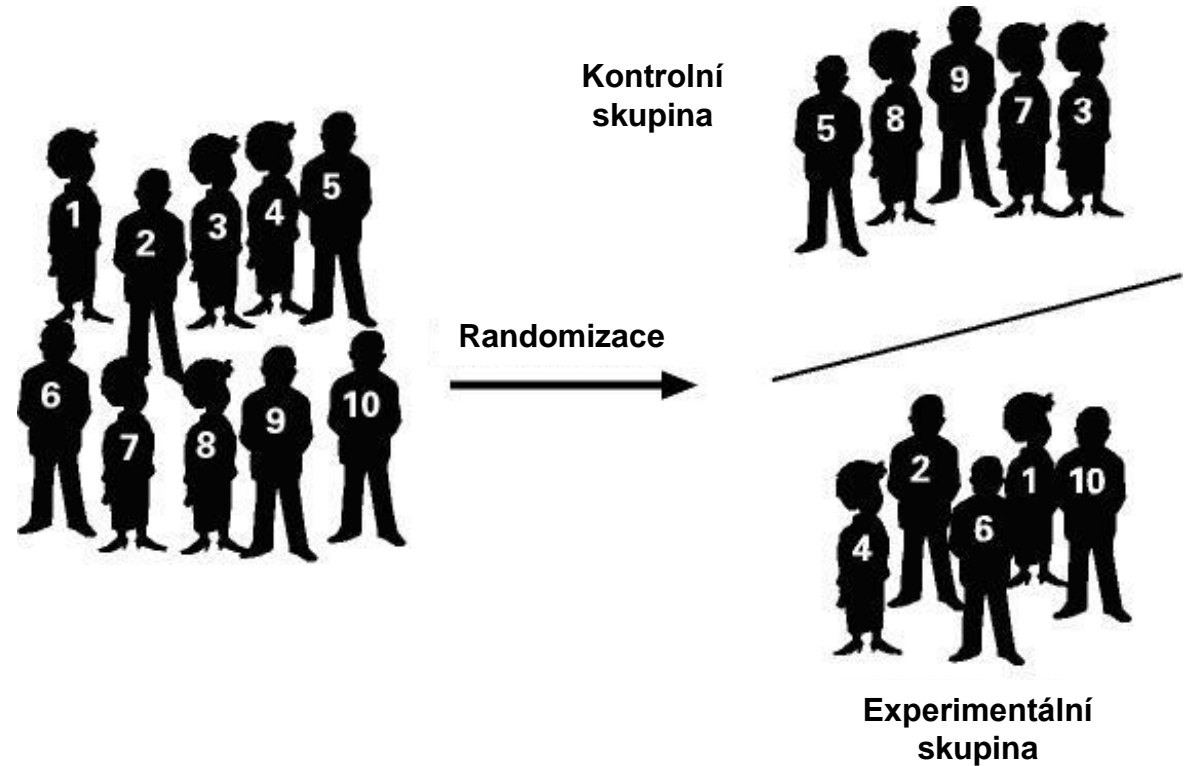
Analyzované populace

- **Intention-to-treat** (ITT) - soubor pacientů plánovaných pro použití intervence bez ohledu na to, zda a v jaké míře byla reálně aplikována.
 - Např. všichni pacienti, kteří podstoupili randomizaci, analyzováno dle léčby přiřazené při randomizaci, bez ohledu na skutečnost
 - „Gold standard” pro fázi III, konzervativní z hlediska superiority (účinek prokázán přesto, že byli započítáni pacienti, u kterých nemohlo k úplnému účinku dojít, odpovídá účinku v reálné populaci)
- **Modified intention-to-treat** (mITT) – podskupina pacientů ITT, kterým byla podána alespoň jedna dávka léčiva.
- **Per-protocol** (PP) - soubor pacientů, kteří odpovídali zařazovacím kritériím a dodrželi protokol (tedy žádné významné odchylky od protokolu).
 - Běžnější pro fázi II (měření “biologického účinku”) a ekvivalenční studie (konzervativní z hlediska ekvivalence; zařazení pacientů neléčených dle protokolu povede ke snížení účinku intervencí, tj. povede k rovnosti)
- **Safety set** - soubor pacientů, kterým byla podána alespoň jedna dávka léčiva
 - Analyzováno na základě léčby skutečně obdržené, bez ohledu na randomizaci

Randomizace a zaslepení studie

Randomizace

Metodika a proces náhodného (pseudonáhodného) rozdělování subjektů hodnocení (pacientů) do dvou nebo více léčebných skupin.

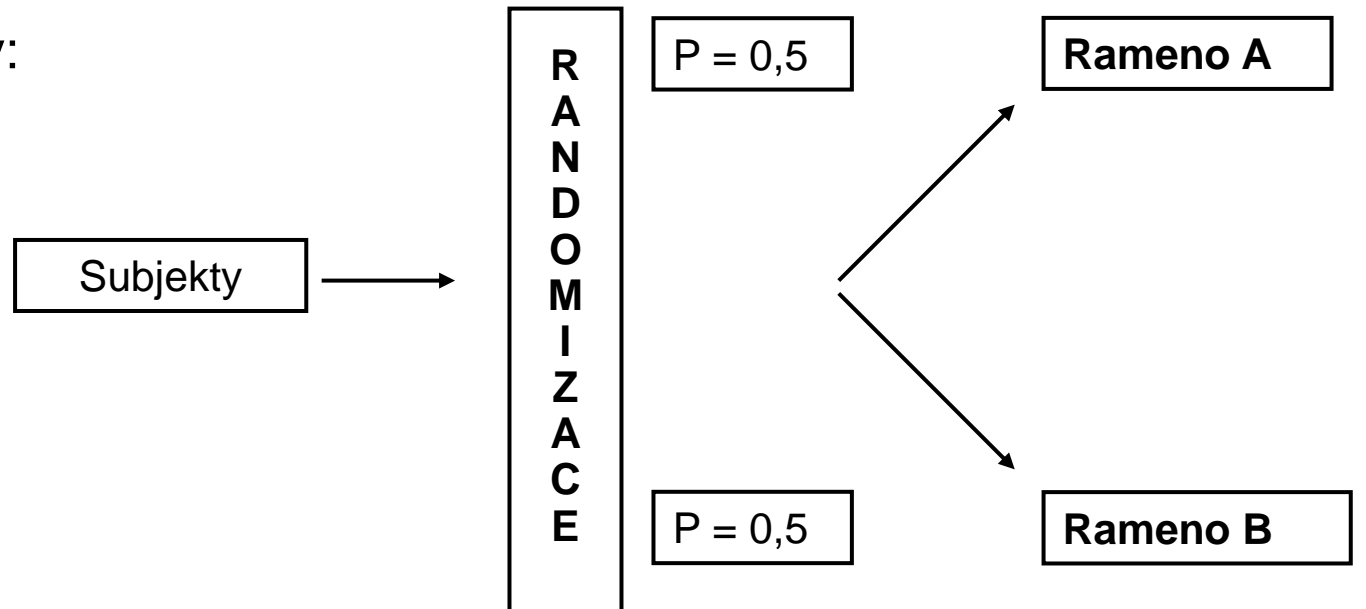


Cíle a smysl randomizační procedury

- Zamezit subjektivnímu a selektivnímu rozdělování pacientů (subjektů) do léčebných skupin a zamezit tak systematické chybě - zkreslení.
- Splnit základní předpoklad validního použití statistických testů - náhodný výběr z populace
- Zabezpečit požadovaný poměr počtu pacientů v jednotlivých léčebných skupinách
- Možnost kontroly rovnoměrné distribuce prognostických faktorů v léčebných skupinách
- Randomizace však není zárukou rovnoměrné distribuce prognostických faktorů – i při minimálním rozdílu v klíčovém parametru mezi jednotlivými skupinami je nutné tento rozdíl brát při statistickém hodnocení v potaz

Kompletní randomizace

- Principem náhodné rozdělování pacientů do definovaných skupin
- Např. klinická studie se dvěma rameny: A,B; N = 100.
- Nevhodná! **Riziko nevyváženého počtu subjektů hodnocení v léčebných skupinách.**
- Při použití kompletní randomizace je pravděpodobnost shodného počtu SH ve dvou léčebných skupinách se 100 SH celkem pouze 8%.



Permutační bloková randomizace

- Princip metody je pravidelné opakování bloků složených z identifikátorů skupin zajišťující stejný počet subjektů v obou skupinách.
- Příklad klinická studie se dvěma rameny: A,B
 1. Požadovaný počet subjektů: $N = 80$
 2. Volba velikosti bloku: 4
 3. Permutace: Celkem 6 možností bloku
 4. Sestavení randomizačního plánu jako sekvence 20 náhodně po sobě jdoucích bloků

1	AABB
2	BBAA
3	ABBA
4	BAAB
5	ABAB
6	BABA

Stratifikovaná permutační bloková randomizace

- Principem je provedení permutační blokové randomizace v rámci definovaných podskupin pacientů.
- Příklad: N = 80 subjektů, 2 ramena (A,B), blok o velikosti 4
 1. Volba stratifikačních kritérií:
 - pohlaví: muž vs. Žena
 - věk: ≤ 50 vs. > 50
 2. Vytvoření čtyř podskupin subjektů na základě všech možných kombinací prognostických faktorů
 3. Bloková randomizace v rámci podskupin
 - Stratifikační faktor musí být parametr zjistitelný před randomizací subjektů, maximální efektivita je při 2-3 stratifikačních parametrech (dle rozsahu studie),
 - Stratifikací nelze kontrolovat všechny prognostické faktory, tento problém je předmětem post-randomizačních vícerozměrných statistických analýz.

Blok	Skupina			
	Muži		Ženy	
	≤ 50	> 50	≤ 50	> 50
1	A	A	B	A
	B	A	B	A
	B	B	A	B
	A	B	A	B
2	B	A	B	A
	B	B	A	B
	A	A	A	A
	A	B	B	B

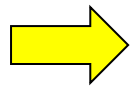
Limitace randomizačních procedur

- **Randomizace nezaručuje externí validitu** klinické studie, ta je dána především vstupními kritérii.
- **Randomizace nezaručuje rovnoměrnou distribuci všech prognostických faktorů** v léčebných skupinách.
- Randomizace je pouze jedna z řady metod využívaných k minimalizaci rizika zkreslení výsledků klinických studií – tzn. není to spása.

Blinding - zaslepení

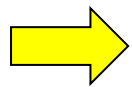
- **Zaslepení studie je proces, kdy pacient a/nebo investigátor nevědí, jakou léčbu pacient dostává** (při paralelním uspořádání studie není znám podaný lék, při cross-over uspořádání studie není známo pořadí podaných léků). Neznamená to, že by pacienti nevěděli, jaké léky jsou ve studii použity!
- **Cílem zaslepení je vyhnout se úmyslnému i neúmyslnému zkreslení výsledků studie.**
- Pacient jinak hodnotí svůj zdravotní stav, když ví, je-li mu podávána nová léčba nebo placebo.
- Lékař může být tímto vědomím také ovlivněn, např. při použití kategoriálního hodnocení.

Blinding - zaslepení



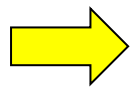
– **Single blinded / jednoduše zaslepená studie**

- pacient neví jaký lék dostává, zatímco lékař jej zná
- téměř se nepoužívají, pouze v případech, kdy nelze jinak (např. v chirurgii)



– **Double blinded / dvojitě zaslepená studie**

- pacient ani lékař neznají léčbu přidělenou pacientovi
- drtivá většina studií je dvojitě zaslepená (např. v onkologii při testování nových modalit léčby)



– **Triple blinded / trojitě zaslepená studie**

- pacient, lékař ani statistik studie neznají přidělenou léčbu
- modifikace dvojitě zaslepené studie
- z hlediska objektivity studie je to další zlepšení

Proč je důležité prospektivně stanovit velikost vzorku?

- Optimalizace velikosti vzorku je **nutná z hlediska korektního statistického zpracování výsledků** a dostatečné síly studie prokázat klinicky zajímavý rozdíl.
- Etické aspekty – nelze zbytečně léčit lidi
- Statistické vlastnosti – při velkém N lze prokázat cokoliv
- Ekonomické aspekty – zbytečné plýtvání prostředky
- Požadavky legislativních autorit (SÚKL, FDA) na výpočet velikosti vzorku
- Rizika neplánovaného počtu subjektů hodnocení:
 - 1. Malý vzorek – ztráta času, nemožnost prokázat rozdíl mezi srovnávanými skupinami pacientů**
 - 2. Velký vzorek – ztráta času a prostředků, průkaz klinicky nevýznamného rozdílu mezi srovnávanými skupinami pacientů**

Jak tedy zní otázka při plánování klinických studií?

- **Kolik subjektů hodnocení potřebujeme** pro korektní ověření klinické hypotézy (klinicky významného rozdílu)?
- Nebudeme-li schopni tyto subjekty z různých důvodů zajistit, jaká je cena za snížení velikosti vzorku? Vede na tzv. power analýzu.
- Důvody pro nedostatečné N:
 - Nedostatek prostředků,
 - Nedostatečná incidence – museli bychom čekat X let na dostatečný počet vhodných pacientů.

Princip testování hypotéz

- Hypotézy jsou testovány na základě testových statistik, které odrážejí, jak moc se náš výsledek liší od počátečních předpokladů = nulové hypotézy.

$$\text{Testová statistika} = \frac{\text{Pozorovaná hodnota} - \text{Očekávaná hodnota}}{\text{Chyba pozorované hodnoty}}$$

N – velikost vzorku

σ – variabilita cílové proměnné

Δ - velikost klinicky významného rozdílu

α, β – pravděpodobnosti falešně pozitivního a negativního výsledku

Testová statistika obsahuje (tj. vychází) z určitých parametrů.

Pokud chceme vypočítat jeden z nich, ostatní musíme znát (expertně odhadnout).

Srovnání dvou výběrů – příklad

- Účinnost dvou preparátů pro léčbu infekce dýchacích cest: primárním cílem studie je odpověď pacienta na podání léku (ano/ne). Standardní přípravek účinkuje u 80% pacientů s tím, že klinicky významný posun by bylo dosažení odpovědi u 90% pacientů, tedy zvýšení účinnosti o 10% při použití nového léku oproti standardu.

Rameno A (standardní lék): účinnost = 80%

Rameno B (nový lék): požadujeme účinnost = 90%

$$\alpha = 0.05$$

$$\beta = 0.20 \rightarrow \text{síla testu: } 1 - \beta = 0.80$$

odhad incidence úspěchů v kontrolní skupině = 80%

klinicky významný rozdíl mezi rameny = 10%

N = 199

pacientů v jednom rameni

pro $\beta = 0.10 \rightarrow$ síla testu: $1 - \beta = 0.90$ je
odhad počtu pacientů v 1 rameni = 266

Interim analýza, subgroup analýzy

Subgroup analýza x interim analýza I

Subgroup analýza:

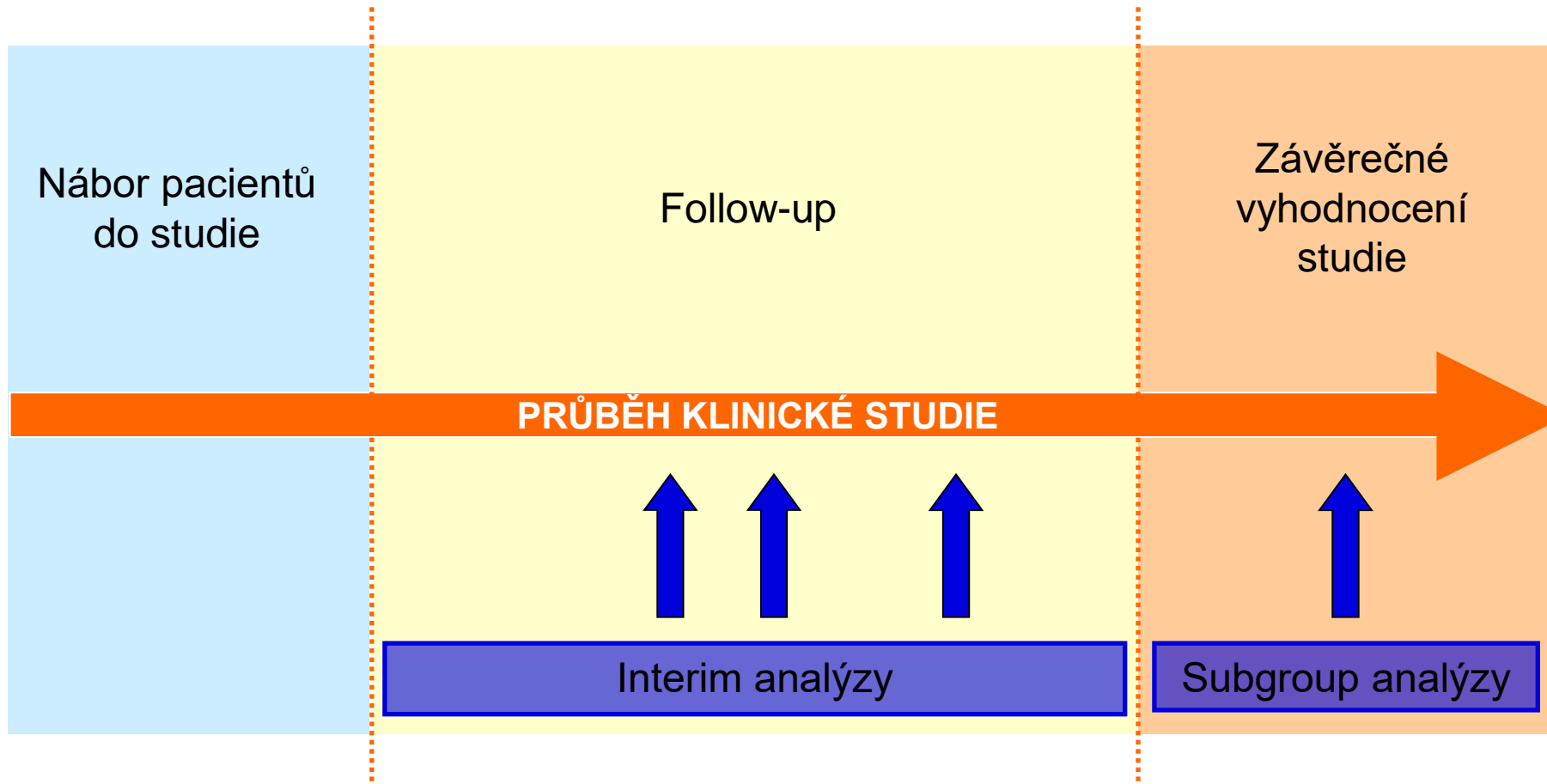
Analýza bezpečnosti a účinnosti léčby prováděná na podskupinách subjektů hodnocení definovaných na základě stanovených vstupních kritérií (jednoho či více).

Interim analýza:

Rozhodnutí o dalším pokračování studie na základě zhodnocení platnosti nulové hypotézy z dostupných dat. Tedy analýza bezpečnosti a účinnosti léčby prováděná před ukončením follow-up.

- V případě interim analýz i subgroup analýz se jedná o násobné testování hypotéz – výsledné p-hodnoty tedy musí být korigovány na počet provedených testů.
- Interim analýza – O'Brien-Fleming kritérium; Subgroup analýza – Bonferroniho korekce.

Subgroup analýza x interim analýza II



Testování hypotéz a možnost chyby

– Při rozhodnutí o výsledku se můžeme splést, rozeznáváme dva druhy chyby v úsudku

Naše rozhodnutí (na základě testu)	(neznámá) realita	
	H_0 platí	H_0 neplatí (e.g. léčba je účinná)
H_0 nezamítáme	správně „nezamítáme“ platnou nulovou hypotézu	chyba II. druhu BETA
H_0 zamítáme (např. účinnost je ukázána)	chyba I. druhu (pravděpodobnost = hladina významnosti) ALFA	správně zamítáme neplatnou nulovou hyp. (pravděpodobnost = síla) 1 – BETA

- Pravděpodobnost alespoň jednoho falešně pozitivního výsledku pro dva testy:

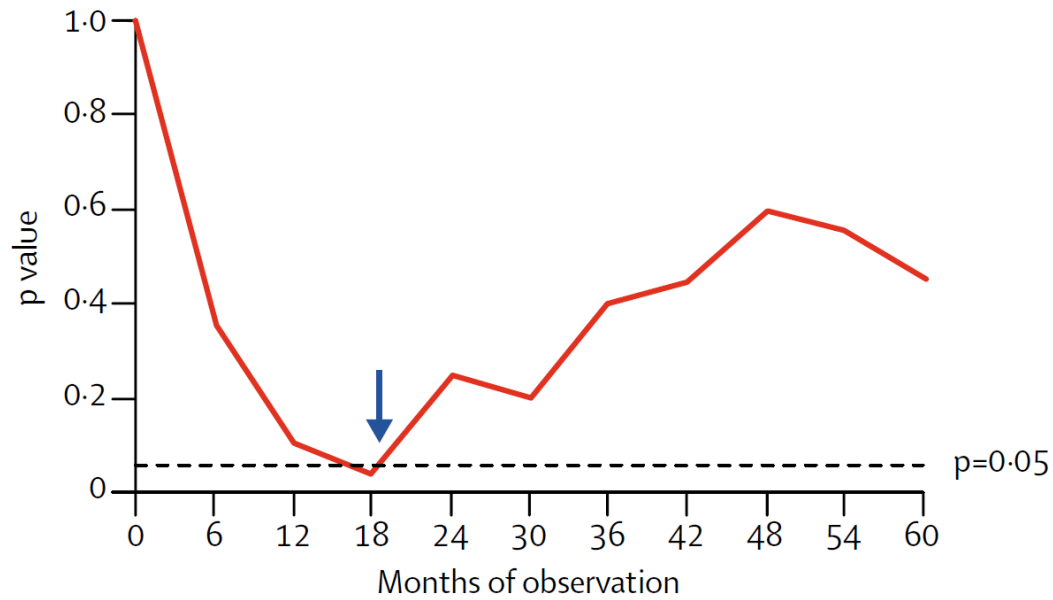
$$1 - 0,95^2 = 0,0975$$

- Pravděpodobnost alespoň jednoho falešně pozitivního výsledku pro deset testů:

$$1 - 0,95^{10} = 0,4013$$

Interim analýza – příklad

- Průběžný monitoring získaných výsledků o účinnosti a bezpečnosti je důležitý! ALE může vést k zavádějícím výsledkům, pokud není proveden korektně.



Kde je chyba?

SCHULZ, Kenneth F.; GRIMES, David A. Multiplicity in randomised trials II: subgroup and interim analyses. *The Lancet*, 2005, 365.9471: 1657-1661.

Interim analýza – řešení

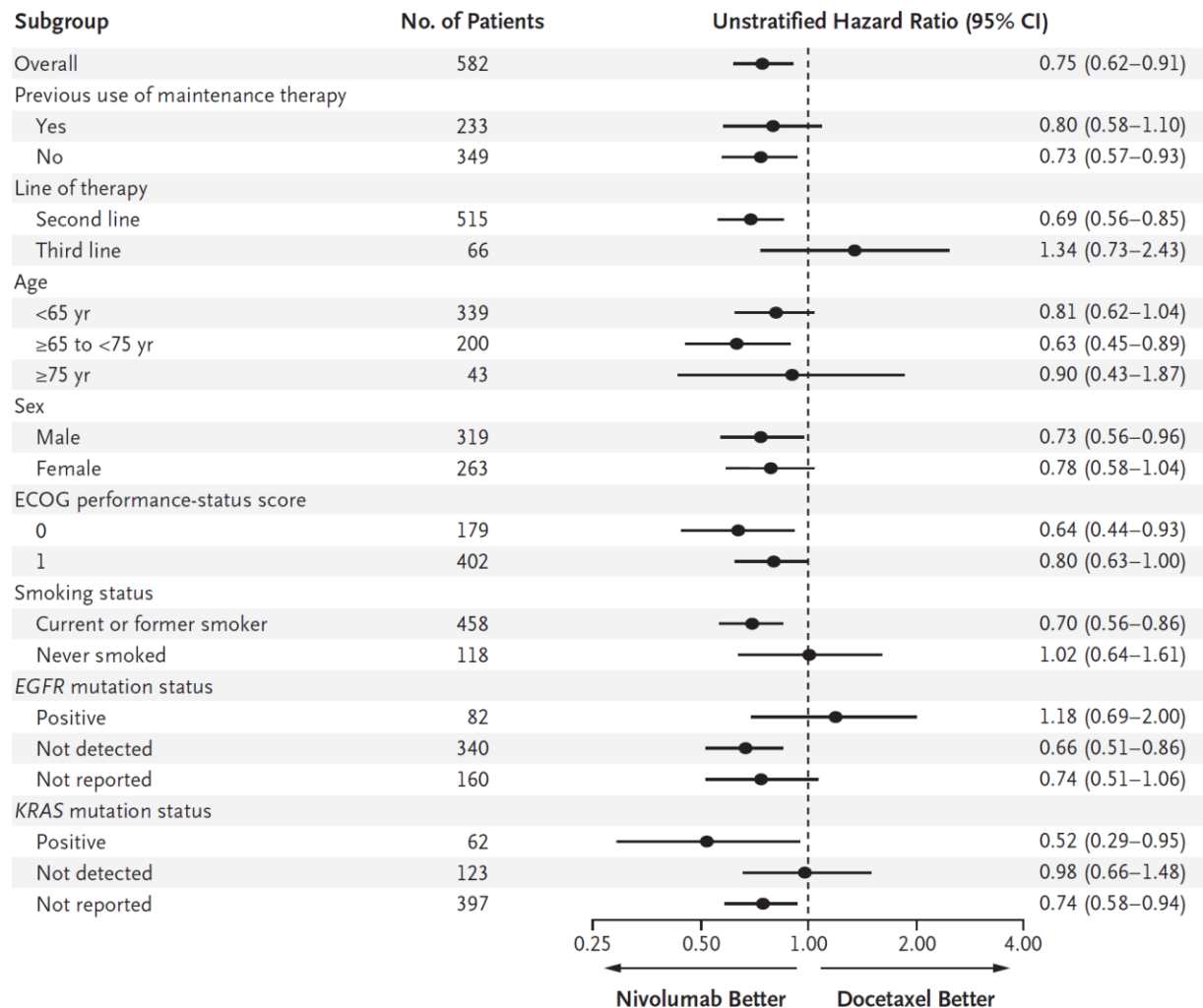
- Hodnocení získaných výsledků o účinnosti a bezpečnosti v rámci interim analýzy na hladině významnosti $\alpha = 0,05$ vede k vyšší pravděpodobnosti falešně pozitivního výsledku.
- Několika autory bylo navrženo řešení v podobě různých hranic pro zamítnutí nulové hypotézy (zastavení studie) s ohledem na počet plánovaných interim analýz.

Number of planned interim analyses	Interim analysis	Pocock	Peto	O'Brien-Fleming
2	1	0.029	0.001	0.005
	2 (final)	0.029	0.05	0.048
3	1	0.022	0.001	0.0005
	2	0.022	0.001	0.014
	3 (final)	0.022	0.05	0.045
4	1	0.018	0.001	0.0001
	2	0.018	0.001	0.004
	3	0.018	0.001	0.019
	4 (final)	0.018	0.05	0.043
5	1	0.016	0.001	0.00001
	2	0.016	0.001	0.0013
	3	0.016	0.001	0.008
	4	0.016	0.001	0.023
	5 (final)	0.016	0.05	0.041

Overall $\alpha=0.05$.

SCHULZ, Kenneth F.; GRIMES, David A. Multiplicity in randomised trials II: subgroup and interim analyses. *The Lancet*, 2005, 365.9471: 1657-1661.

Subgroup analysis – příklad



- Výhodou je možnost nalezení skupiny pacientů, u kterých je/není léčba účinnější a bezpečnější – cílená terapie.
- Na menších podsouborech nemá zvolený test dostatečnou sílu pro prokázání rozdílů v účinnosti/bezpečnosti.
- **Zvýšení rizika falešně pozitivního výsledku s ohledem na násobné testování.**

Borghaei et al., 2015. Nivolumab versus Docetaxel in Advanced Non-squamous Non-small Cell Lung Cancer. *N Engl J Med.* 2015 Oct 22; 373(17): 1627–1639.

Subgroup analýzy

- Počet plánovaných subgroup analýz by měl být uveden v protokolu studie (tedy před jejím zahájením).
- Pokud je to možné, před zahájením studie by měl být navržen optimální počet pacientů také pro plánované subgroup analýzy.
- Hladina významnosti pro testování hypotéz by v případě více testů měla být korigována dle jejich počtu.
- **Výsledky subgroup analýz nelze přeceňovat – v kontextu klinické studie je hlavním výsledkem studie vyhodnocení primárního endpointu studie. Na základě výsledku subgroup analýz nelze vytvářet klinická doporučení.** Výsledky subgroup analýz by měly být interpretovány jen jako hypotézy pro další výzkum.

Legislativa

Etika v klinickém výzkumu

- Cíle nastavení etických požadavků na klinický výzkum:
 1. Minimalizovat možnost zneužití subjektů
 2. Zajistit práva a prospěch subjektů

Významné dokumenty

- Hippokratova přísaha - Ženevská deklaráce 1948
- **Norimberský kodex** – 1947
- Všeobecná deklaráce lidských práv – OSN, 1948
- **Helsinská deklaráce** Světové lékařské asociace, 2013 – tato deklaráce se týká přímo lékařského výzkumu s účastí lidských subjektů, první verze již v r. 1964
- Belmontská zpráva – 1979
- Úmluva na ochranu lidských práv a důstojnosti lidské bytosti v souvislosti s aplikací biologie a medicíny – Rada Evropy, 1997 – viz Sbírka mezinárodních smluv ČR č. 96/2001; Mezinárodní směrnice pro biomedicínský výzkum

ICH GCP

- **Mezinárodní konference o harmonizaci (ICH)** technických požadavků, jejímž úkolem je registrace léčivých přípravků pro humánní použití.
- ICH přijala v roce 1996 principy správného provádění klinických studií = **GCP (Správná klinická praxe)**.
- Omezená působnost (USA + ES + Japonsko: lékové úřady a farmaceutický průmysl)
- **Soubor psaných standardů pro plánování, provádění, vedení, monitorování, auditování (kontrolu), zapisování, analyzování a hlášení klinických hodnocení**, který zajišťuje, aby údaje a výsledky z klinických hodnocení prováděných v souladu se správnou klinickou praxí byly věrohodné, přesné a aby byla zajištěna ochrana práv a integrity subjektů hodnocení i důvěrnost údajů o subjektech hodnocení.

ICH GCP – cíl

- Poskytnout **jednotný standard** pro Evropské společenství, Japonsko a USA, aby se usnadnilo vzájemné přijímání klinických dat zodpovědnými orgány.
- **Usnadnit celosvětový vývoj léčiv** pomocí harmonizace technických požadavků pro registrace léků pro použití u lidí.
- Helsinská deklarace vs. ICH GCP:
 - Helsinská deklarace zaměřena na morální principy
 - ICH GCP zaměřeno na postupy

M U N I
M E D

**Děkuji za
pozornost**